

## 2940.0 - Census of Population and Housing: Details of Overcount and Undercount, Australia, 2016

Latest ISSUE Released at 11:30 AM (CANBERRA TIME) 23/02/2018

## Summary

### Key Findings

#### KEY FINDINGS

This publication presents estimates of undercount and overcount for the 2016 Census of Population and Housing (Census), resulting from the 2016 Census Post Enumeration Survey (PES).

The Census net undercount is the difference between the PES estimate of how many people should have been counted in the Census and the actual Census count (including imputed persons).

#### Australia

- The Census net undercount rate in 2016 was 1.0% (equivalent to 226,407 persons)
- The net undercount rate was lower in 2016 compared with 2011 (1.0% and 1.7%, respectively)
- The Contact sector showed a net undercount of 875,915 persons, indicating that the number of persons missing from responding Census dwellings (1,150,588 persons) exceeded the number of persons counted multiple times or in error (274,673 person)
- The Non-contact sector had a net undercount of -649,509 persons (i.e. a net overcount), which is essentially a measure of over-imputation for non-responding dwellings in the Census that were deemed occupied.

#### States and Territories

- The Northern Territory recorded the highest net undercount rate of all states and territories (5.0%), while the Australian Capital Territory recorded the lowest net undercount rate (-1.1%; i.e. a net overcount)
- Victoria was the only state or territory to exhibit a higher net undercount rate in 2016 (1.4%) compared with 2011 (1.1%).

#### Key Population Characteristics

- The highest net undercount rate for age groups was for the 0-4 year olds (5.1%), followed by the 20-24 and 25-29 year age groups (5.0% and 4.9%, respectively)
- Males were more likely to be missed in the Census compared with females, with net undercount rates of 1.5% and 0.4%, respectively
- The net undercount rate for Aboriginal and Torres Strait Islander peoples was 17.5% (equivalent to 137,750 persons). This is only slightly higher than 2011 (17.2%)
- For Country of birth, persons born in Australia had the highest net undercount rate (8.1%) while China showed the largest change in net undercount of the ten highest ranked countries (from 14.9% in 2011 to 6.2% in 2016).

## About the Census Post Enumeration Survey

### ABOUT THE CENSUS POST ENUMERATION SURVEY

The Census Post Enumeration Survey (PES) is a household survey conducted by the Australian Bureau of Statistics (ABS) shortly after each Census of Population and Housing (Census), in order to provide an independent measure of Census coverage. This release presents estimates of undercount and overcount for the 2016 Census, resulting from the 2016 PES. This publication also details how the 2016 PES was conducted and how 2016 PES estimates have been calculated.

The 2016 PES included people from approximately 42,000 fully responding households across Australia. Information was collected for everyone present in the household. In addition to obtaining basic demographic information, questions were asked about each person's usual residence, their location on Census night, and any other addresses where they might have been counted in the Census.

The PES results are used to determine how many people should have been counted in the Census, how many people were missed (undercount) and how many were counted more than once or in error (overcount). The PES also measures the level of imputation error in Census processing (where imputation is used to assign values for persons in non-responding Census dwellings). Census **net undercount** is the difference between the PES estimate of how many people should have been counted in the Census and the actual Census count (including imputed persons in non-responding dwellings).

PES estimates of net undercount augment the Census counts for the purpose of deriving resident population estimates for Australia and its states and territories.

The Census counts presented in this publication may differ slightly from aggregate counts in other Census products as the counts used for PES estimation relate to usual residents of Australia and exclude those who reside in Other Territories (e.g. Norfolk Island or Antarctica). The Census counts in this publication include persons who were imputed into non-responding dwellings, as well as persons who responded to the Census.

## ROUNDING

Estimates in this publication have been rounded. Proportions presented in this publication are based on unrounded estimates. Calculations using rounded estimates may differ from those published.

# Coverage in the Census

## COVERAGE IN THE CENSUS

Tuesday 9 August 2016 was Census night in Australia. All people present in Australia on this night, with the exception of foreign diplomats and their families, should have been included on a Census form at the place where they stayed.

The Census of Population and Housing (Census) is the largest statistical collection undertaken by the Australian Bureau of Statistics (ABS) and one of the most important. Its objective is to accurately measure the number of people in Australia on Census night, their characteristics, and the dwellings in which they live. Due to its size and the complexity of this task, however, it is inevitable that some people will be missed from the Census and some will be counted more than once.

The Census Post Enumeration Survey (PES) is run shortly after each Census as a way to independently measure Census coverage. The PES determines how many people should have been counted in the Census, how many were missed, how many were counted more than once, and how many were counted in error.

The PES also provides an estimate of Census imputation error; i.e. the difference between the number of people imputed into non-responding dwellings during Census processing and the number of people who should have been counted in those dwellings. It also provides information on the characteristics of those in the population who have been missed or overcounted, including an indication of those characteristics which may have changed or been misclassified between the Census and PES.

Some of the reasons why people may have been missed in the Census (i.e. undercounted) include:

- they were travelling and were difficult to contact
- they mistakenly thought they were counted elsewhere
- there was insufficient space on the Census form in the household where they were staying and they did not obtain additional forms
- the person completing the form thought that certain people (e.g. young babies, the elderly or visitors) should not be included
- they did not wish to be included due to concerns about confidentiality or a more general reluctance to participate
- the dwelling in which they were located was missed because it was difficult to find (e.g. in a remote or non-residential area) or the dwelling was not included on the ABS Address Register
- a Census form (including online registration) was not supplied as the dwelling was mistakenly recorded as unoccupied, and no forms for that dwelling were requested.

Some of the reasons why people may have been counted more than once or in error (i.e. overcounted) include:

- they were included on the Census form at the dwelling where they usually live, even though they stayed and were counted elsewhere on Census night
- they moved during the Census period and completed forms at both their previous and new address
- they were overseas on Census night and so should not have been counted at all, but were included on the Census form at the dwelling where they usually live.

# What is Net Undercount?

## WHAT IS NET UNDERCOUNT?

While every effort is made to eliminate these potential causes of error, some undercount and overcount will inevitably occur. Net undercount for any category of person is the difference between the PES population estimate (i.e. estimate of the number of people who **should** have been counted in the Census) and the **actual** Census count.

The Census count includes:

- all responding persons who were counted on at least one Census form, with multiple counts for persons who were included on more than one form
- persons who have been imputed into non-responding Census private dwellings during data processing
- imputed records for persons who spent Census night at a non-private dwelling (e.g. a hotel or caravan park) but did not complete a Census form there.

Net undercount (or overcount) is therefore a measure of the combined outcome of Census enumeration and Census imputation.

For more information about imputed persons for non-responding Census dwellings and the adjustments made for them in the PES estimates, see **Components of Net Undercount** on the Summary tab.

Net undercount is also presented in this publication as a rate. The rate is the net undercount (or overcount) as a percentage of the PES estimate for a given population (i.e. as a percentage of the number of people who **should** have been counted in the Census). The Census is typically associated with a net undercount; however, rates of net undercount or overcount can vary significantly for different population groups depending on factors such as sex, age, Indigenous status, and geographic location, and on whether these characteristics have changed or been misclassified between the Census and PES.

## Key Uses of Net Undercount

### KEY USES OF NET UNDERCOUNT

Undercount and overcount are the primary measure of Census coverage. They provide users with an assessment of the completeness of Census counts and a measure of Census imputation performance. As such, undercount and overcount can be used to evaluate the effectiveness of Census collection procedures and data processing, so improvements can be made for future Censuses.

Net undercount is a key component in rebasing the Estimated Resident Population (ERP) for 30 June of the Census year. Rebasing is done by adding the net undercount from PES to the new Census population counts, before applying further demographic adjustments in order to recalibrate ERP to the latest Census.

Accurate estimates of the resident population are required for a wide range of uses including:

- the allocation of seats to states and territories in the House of Representatives of the Australian Parliament
- the distribution of Commonwealth payments to states and territories
- demographic, social and economic research
- policy research, evaluation and government reporting.

For more information on the calculation of ERP for 30 June 2016, based on results from the 2016 Census and PES, see the ABS publication *Australian Demographic Statistics*, December 2016 (cat. no. 3101.0), released on 27 June 2017.

## Estimates of Net Undercount

### ESTIMATES OF NET UNDERCOUNT

Estimates in the following sections will be presented either as the total net undercount for a category of person (i.e. the difference between the PES population estimate and the Census count) or as the net undercount rate. For further information on net undercount and the net undercount rate, see **What is Net Undercount?** on the Summary tab.

The Census counts people on a place of enumeration basis (i.e. where they were on Census night) and a place of usual residence basis. All estimates of net undercount based on geography in this publication have been calculated on a place of usual residence basis.

## Australia

### AUSTRALIA

The 2016 Census counted 23,397,296 usual residents of Australia, who were in the country on Census night (including 1,183,519 persons who were imputed into non-responding dwellings). This was 226,407 persons fewer than the PES estimate of the usual resident population who were present in Australia on Census night. This equates to a net undercount rate of 1.0%.

The net undercount rate is lower in 2016 compared with 2011 (1.7%). The 2016 estimate continues the historical trend of almost complete coverage in Australian Census counts.

**Net Undercount Rate, Australia, 1971-2016**

	1971 %	1976 %	1981 %	1986 %	1991 %	1996 %	2001 %	2006 %	2011 %	2016 %
Net undercount	1.4	2.7	1.9	1.9	1.8	1.6	1.8	2.7	1.7	1.0
Standard Error (SE)	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2

It is important to note that the PES is designed to provide the best measure of Census coverage at a single point in time rather than as a time series. PES estimates of net undercount are not strictly comparable over time due to changes in both Census and PES methodologies. For example:

- a new estimation process was introduced in the 2006 PES, along with the inclusion of remote areas and Discrete Communities in the PES sample
- automated data linking between PES and Census person records was introduced in 2011
- the 2016 Census underwent a number of changes to its collection methodology, with some changes to PES enumeration procedures as a result.

## Age and Sex

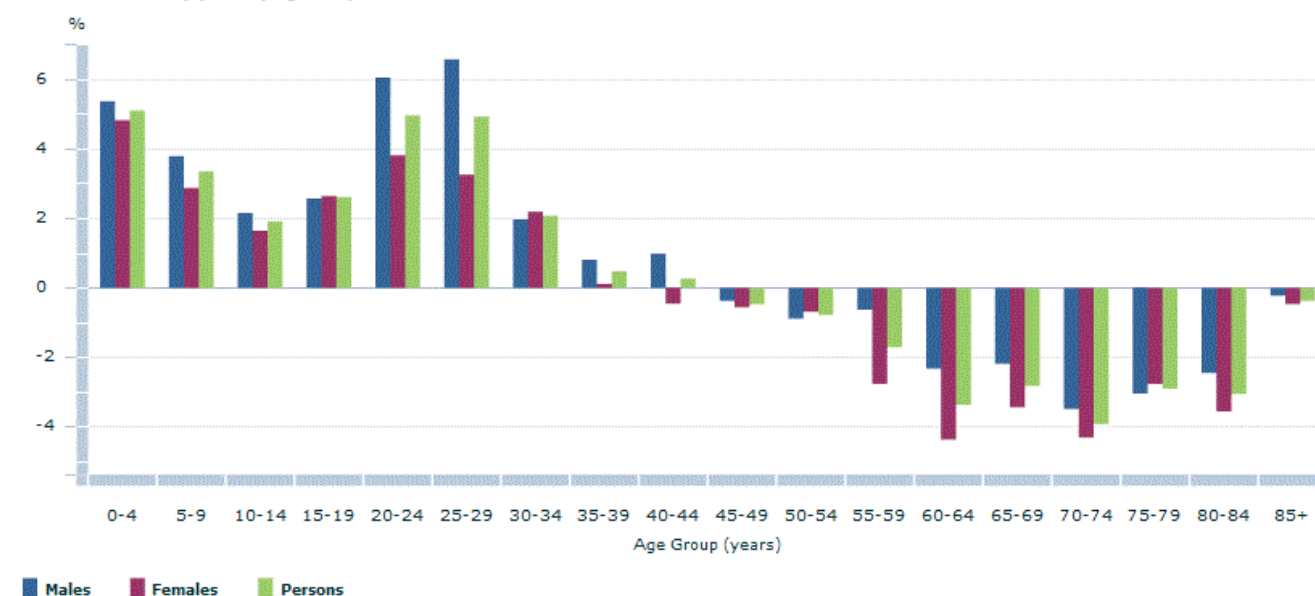
### AGE AND SEX

Previous Censuses (both in Australia and overseas) have observed that young adults aged 20-29 years are more likely to be missed, while older adults are more likely to be counted or overcounted in a Census and young children are often mistakenly omitted from Census forms. Males are also traditionally more likely to be missed than females.

These trends hold true for the 2016 Census.

The highest net undercount rate in 2016 was for the 0-4 year olds (5.1%), followed by the 20-24 and 25-29 year age groups (5.0% and 4.9%, respectively). This is a change from 2011, where the young adults had the highest net undercount rates. In contrast, the older age groups all showed low net undercount rates, with the lowest rate for people aged 70-74 years (-3.9%; i.e. a net overcount). Males were also more likely to be missed in the Census compared with females, with net undercount rates of 1.5% and 0.4%, respectively.

Net Undercount Rate(a), Sex by Age Group, 2016



Save Chart Image

Australian Bureau of Statistics

© Commonwealth of Australia 2017.

**Footnote(s):** (a) A negative value indicates a net overcount.

**Source(s):** Census of Population and Housing: Details of Overcount and Undercount, Australia - 2016

For males, those aged 25-29 years had the highest net undercount rate of all age groups (6.6%) followed by 20-24 year olds (6.1%). These age groups also had the highest net undercount rates in 2011.

For females, the 0-4 year olds had the highest net undercount rate (4.8%), followed by the 20-24 year olds (3.8%) and the 25-29 year olds (3.3%). The net undercount rate for females aged 60-64 years was the lowest of all age groups (-4.4%), although a net overcount can be seen for all female age groups 40 years and over.

Children showed a higher net undercount rate in 2016. Compared with 2011, the net undercount rates were higher for 0-4 year olds (from 1.2% to 5.1%), for 5-9 year olds (from 1.5% to 3.4%), and for 10-14 year olds (from 0.4% to 1.9%).

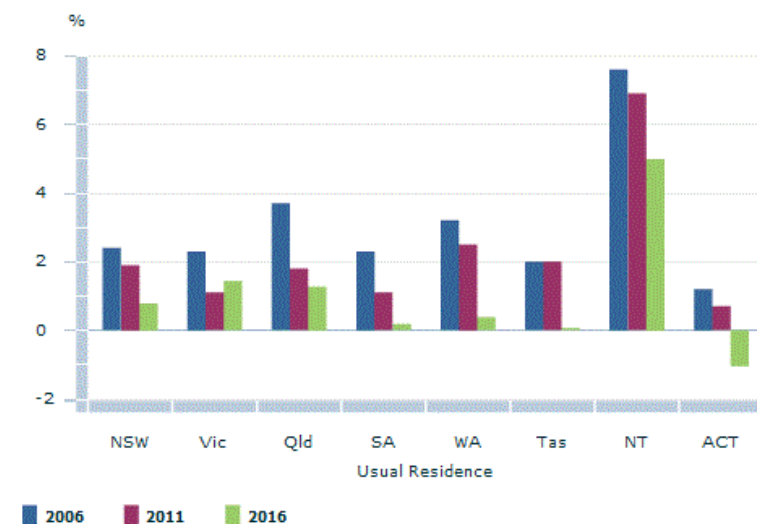
Care should be taken when comparing the 2016 and 2011 net undercount estimates for age, due to an inconsistency in the calculation of age for estimation purposes. The 2016 PES collected the respondent's age as at the time of the PES interview and this was (correctly) backdated to 9 August 2016 (i.e. their age on Census night) using date of birth, where provided. In 2011, however, age was not backdated to Census night for estimation purposes. This would have produced a small downward bias to the 2011 PES population estimate (and hence to the net undercount rate) for 0-4 year olds and a small upward bias for the 85 and over age group.

## State and Territories

### STATES AND TERRITORIES

In 2016, the Northern Territory recorded the highest net undercount rate of all states and territories (5.0%), while the Australian Capital Territory (ACT) recorded the lowest net undercount rate (-1.1%). This is consistent with previous PESs, although the ACT had a net overcount for the first time.

Net Undercount Rate(a), State & Territory of Usual Residence, 2006-2016



[Save Chart Image](#)

Australian Bureau of Statistics

© Commonwealth of Australia 2017.

**Footnote(s):** (a) A negative value indicates a net overcount.

**Source(s):** Census of Population and Housing: Details of Overcount and Undercount, Australia - 2016

While the two territories continue to reflect the minimum and maximum net undercount rates, trends across the states have shifted for 2016. The largest changes in net undercount in 2016 were for Western Australia (2.5% in 2011 to 0.4% in 2016) and Tasmania (2.0% in 2011 to 0.1% in 2016). Along with South Australia (0.2%), they now have the lowest net undercount rates across the six states. In contrast, Victoria was the only state or territory to exhibit a higher net undercount rate in 2016 (1.4%, compared with 1.1% in 2011).

## Greater Capital City and Rest of State Region

### GREATER CAPITAL CITY AND REST OF STATE REGION

Different issues are encountered in enumerating different areas of Australia and these are reflected in the net undercount rates. In urban areas, locating dwellings is generally easier, but contacting occupants and ensuring their participation can be more difficult. In contrast, in rural and remote areas where dwellings may be scattered over a wider area, locating the dwellings can cause considerable difficulties.

In 2016, South Australia, Tasmania and the Northern Territory all had higher net undercount rates in their rest of state regions compared with their greater capital cities. All other states and territories had a lower net undercount rate in their rest of state regions (compared with their greater capital cities).

At the national level, the net undercount rate in 2016 was lower for the rest of state regions (0.5%) compared with greater capital cities (1.2%). The difference was less in 2011 (1.7% and 1.8%, respectively). As was the case in both 2011 and 2006, the Northern Territory continues to show the largest difference in net undercount rates between its greater capital city (3.5%) and rest of state region (7.2%).

Net Undercount Rate(a), State/Territory of Usual Residence by Region, 2016

	Greater capital city		Rest of state		Total	
	%	SE	%	SE	%	SE
New South Wales	1.0	0.5	0.4	0.7	0.8	0.4

Victoria	2.0	0.5	-0.3	0.8	1.4	0.4
Queensland	1.9	0.7	0.6	0.7	1.3	0.5
South Australia	-0.5	0.5	2.4	1.6	0.2	0.5
Western Australia	0.5	0.6	-0.2	1.8	0.4	0.6
Tasmania	-0.2	1.3	0.3	1.0	0.1	0.7
Northern Territory	3.5	1.5	7.2	3.0	5.0	1.5
Australian Capital Territory	-1.1	1.4	—	—	-1.1	1.4
<b>Australia</b>	<b>1.2</b>	<b>0.3</b>	<b>0.5</b>	<b>0.4</b>	<b>1.0</b>	<b>0.2</b>

— Nil or rounded to zero (including null cells).

(a) A negative value indicates a net overcount.

## Registered Marital Status

### REGISTERED MARITAL STATUS

In 2016, the net undercount rate was highest for people identified as never married (3.9%) and lowest for people widowed, divorced or separated (-4.5%). This is consistent with 2011, although the net overcount rate in the widowed, divorced or separated group was much higher in 2016 compared with 2011.

#### Net Undercount Rate(a), Registered Marital Status, 2006-2016

	2006		2011		2016	
	%	SE	%	SE	%	SE
Never married(b)	4.6	0.3	3.7	0.3	3.9	0.3
Widowed, divorced or separated	1.2	0.5	-0.8	0.5	-4.5	0.6
Married	1.0	0.2	0.2	0.2	-0.9	0.3

(a) A negative value indicates a net overcount.

(b) Includes those who are living with a de facto partner and have never been in a registered marriage.

It is important to consider the strong relationship with age when interpreting net undercount estimates by registered marital status. For example, the overcount in the older age groups, coupled with Census imputation performance for this age group, is a likely driver of the increasing overcount in the widowed, divorced or separated population.

## Indigenous Status

### INDIGENOUS STATUS

Special procedures are used in the Census to support the enumeration of the Aboriginal and Torres Strait Islander population, particularly in remote areas, as counting this population continues to present a number of challenges.

The 2016 PES estimated that 786,689 Aboriginal and Torres Strait Islander peoples should have been counted in the Census, compared with 648,939 persons who were counted. This is equivalent to a net undercount of 137,750 persons, or a rate of 17.5%. This is only slightly higher than 2011, which estimated a net undercount of 114,188 persons, or a rate of 17.2%.

#### Net Undercount Rate, Indigenous Status(a), 2011-2016

	PES population estimate no.	SE	Census count(b)(c) no.	Net undercount no.	SE	Net undercount rate %	SE
2016							
Indigenous	786 689	19 776	648 939	137 750	19 776	17.5	2.1
Non-Indigenous	22 837 014	46 483	21 337 326	1 499 688	46 483	6.6	0.2
Not stated	—	—	1 411 031	—	—	—	—
2011							
Indigenous	662 335	14 274	548 147	114 188	14 274	17.2	1.8
Non-Indigenous	21 216 926	37 272	19 898 127	1 318 799	37 272	6.2	0.2
Not stated	—	—	1 058 447	—	—	—	—

— Nil or rounded to zero (including null cells).

(a) Net undercount is based on Census counts for a category. In the Census, Indigenous status was set to not stated where the response was blank or where imputed person records were created for non-responding dwellings. Hence components of undercount for Indigenous status do not sum to the Australia total.

(b) Includes imputed persons in non-responding dwellings. These were all given an Indigenous status of not stated.

(c) Refers to Census counts which correspond to the scope of the PES and may differ slightly from aggregate counts in other Census products.

It is important to note that Indigenous status, as collected in both the Census and PES, is based on responses to a question related to information that some people will consider personal and sensitive. Respondents may choose to indicate in the Census that they are of Aboriginal or Torres Strait Islander origin, not of Aboriginal or Torres Strait Islander origin, or they may choose to not answer the question at all.

If no answer is provided, the Census does not impute for this missing response (which is also the case for imputed persons). While the person (real or imputed) will continue to be counted in broad-level Census counts, they will not be included in the Census counts for Indigenous status. There were 1,411,031 people (6.0%) whose Indigenous status was not stated in the 2016

Census, compared with 1,058,447 people (4.9%) in 2011.

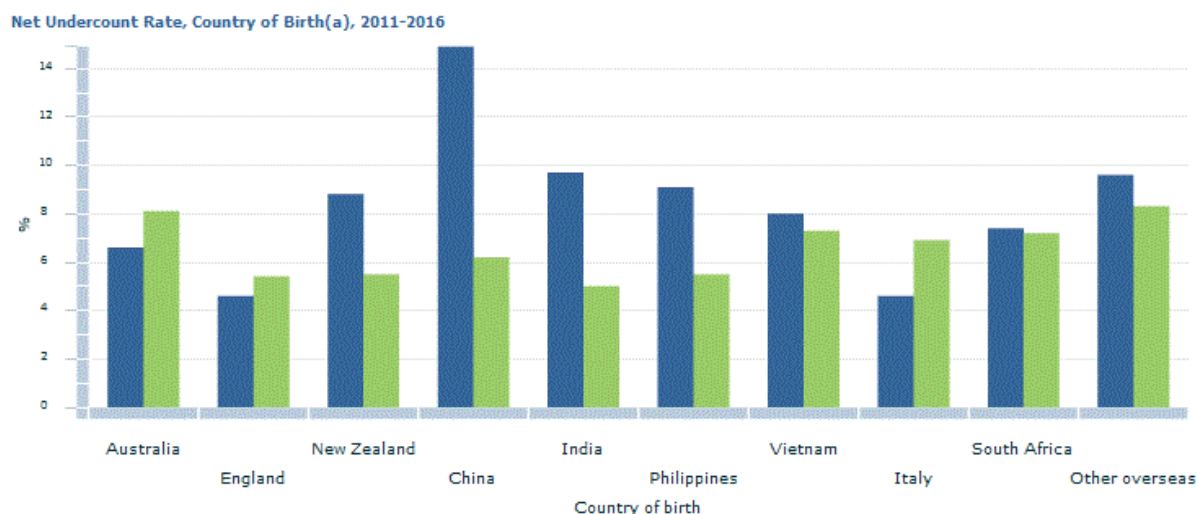
For information on how Census not stated responses feed into estimates of net undercount, see **Components of Net Undercount** on the Summary tab.

## Country of Birth

### COUNTRY OF BIRTH

The net undercount rates for Country of birth are displayed for the ten highest ranked countries (in terms of population residing in Australia) according to the 2016 Census.

Of those countries listed, persons born in Australia had the highest net undercount rate (8.1%), followed by Vietnam (7.3%) and South Africa (7.2%). This is a change from 2011, where Australia-born persons had the seventh highest net undercount, behind China and India as the top two. In 2016, persons born in India had the lowest net undercount rate (5.0%) followed by those born in England (5.4%) and the Philippines (5.5%). China showed the largest change in net undercount of the ten highest ranked countries, with a difference of 8.7 percentage points (from 14.9% in 2011 to 6.2% in 2016).



■ 2011 ■ 2016

Save Chart Image

Australian Bureau of Statistics

© Commonwealth of Australia 2017.

**Footnote(s):** (a) Malaysia replaced Scotland in the ten highest ranked countries in 2016. Both countries have been included in 'Other overseas' to enable a valid comparison between 2011 and 2016.

**Source(s):** Census of Population and Housing: Details of Overcount and Undercount, Australia - 2016

It is important to note that people who have come to Australia from other countries and whose first language is not English may find completing a Census form more difficult than other Australians. For several Censuses, special strategies have been employed to promote an understanding of the Census among migrants and to provide assistance in a range of languages.

Similar to Indigenous status, Country of birth may also be left blank in the Census. Since Census does not impute a Country of birth for these people, their Country of birth is set to not stated in Census counts (which is also the case for imputed persons). However, they contribute to net undercount estimates based on the category in which they should have been counted, as reported in the PES. There were 1,622,118 people (6.9%) whose Country of birth was not stated in the 2016 Census, compared with 1,195,432 people (5.6%) in 2011.

For information on how Census not stated responses feed into estimates of net undercount, see **Components of Net Undercount** on the Summary tab.

## Undercount Adjustment Factors

### UNDERCOUNT ADJUSTMENT FACTORS

While estimates of net undercount are important for an effective understanding of the completeness of Census counts, undercount adjustment factors provide an indication of how much the Census count for a given category would need to be adjusted in order to reflect the PES population estimate for that category.

The undercount adjustment factor is the ratio of the PES population estimate to the actual Census count. This factor can be applied to the Census count for any category to indicate how many people should have been counted in that category. For example, the Census count of 23,397,296 persons in Australia multiplied by the adjustment factor of 1.010 (or unrounded: 1.00967662) indicates that 23,623,703 persons should have been counted in the Census.

#### Population Estimates, Census Counts and Undercount Adjustment Factors, 2016

	PES population estimate		Census count (a)(b) no.	Net undercount		Undercount adjustment factor	
	no.	SE		no.	SE	no.	SE
New South Wales	7 539 414	30 837	7 480 220	59 194	30 837	1.008	0.004
Victoria	6 012 621	25 059	5 926 593	85 028	25 059	1.015	0.004
Queensland	4 763 760	22 882	4 703 210	60 550	22 882	1.013	0.005
South Australia	1 679 607	8 595	1 676 633	2 974	8 595	1.002	0.005
Western Australia	2 483 901	14 081	2 474 424	9 477	14 081	1.004	0.006
Tasmania	510 261	3 444	509 965	296	3 444	1.001	0.007
Northern Territory	240 874	3 720	228 854	12 020	3 720	1.053	0.016
Australian Capital Territory	393 265	5 347	397 397	-4 132	5 347	0.990	0.013
<b>Australia</b>	<b>23 623 703</b>	<b>48 328</b>	<b>23 397 296</b>	<b>226 407</b>	<b>48 328</b>	<b>1.010</b>	<b>0.002</b>

(a) Includes imputed persons in non-responding dwellings.

(b) Refers to Census counts which correspond to the scope of the PES and may differ slightly from aggregate counts in other Census products.

The undercount adjustment factor is not, and therefore should not be, used alone to derive an alternative measure of the Estimated Resident Population (ERP). Official population estimates include additional data and adjustments for usual residents of Australia, such as for those who were temporarily overseas on Census night. For information on the calculation of ERP, see the ABS publication Australian Demographic Statistics, December 2016 (cat. no. 3101.0), released on 27 June 2017.

## Components of Net Undercount

### COMPONENTS OF NET UNDERCOUNT

While net undercount estimates measure the completeness of Census counts, and adjustment factors can be used to adjust Census counts, it is also useful to understand the contributing components of net undercount.

For the purposes of PES estimation, persons are categorised into the contact sector or the non-contact sector.

The **contact sector** includes:

- persons in dwellings for which a Census form was received before the commencement of PES enumeration (which includes persons overcounted and persons missed from these forms)
- persons from occupied dwellings that were entirely missed by the Census
- persons missed by Census because their dwellings were mistakenly deemed unoccupied on Census night.

The **non-contact sector** includes:

- persons in dwellings deemed occupied on Census night, from which no Census form was received. Data for these persons were imputed during Census processing
- persons in dwellings whose Census form was received after the commencement of PES enumeration (i.e. a late return)
- persons with insufficient personal identifier information on their Census form.

The PES traditionally measures a net undercount of persons in the contact sector, with the number of persons missing from Census forms (including persons whose dwelling was missed entirely) generally exceeding the number of persons counted multiple times or in error. In contrast, the non-contact sector is typically characterised by an overcount of persons, which is essentially a measure of over-imputation for non-responding dwellings in the Census that were deemed occupied.

These trends hold true for the 2016 PES.

### Components of Net Undercount, 2011-2016

		Contact sector			Non-contact sector	Total
		Gross undercount(a)	Gross overcount(b)	Net undercount	Net undercount(c)(d)	Net undercount
<b>ESTIMATE (no.)</b>	2016	1 150 588	274 673	875 915	-649 509	226 407
	2011	885 380	144 341	741 039	-366 499	374 540
	Change(%)	+30.0	+90.3	+18.2	+77.2(e)	-39.6
<b>PERCENT (%) (f)</b>	2016	4.9	1.2	3.7	-2.7	1.0
	2011	4.0	0.7	3.4	-1.7	1.7
	Change(%pts.)	+0.9	+0.5	+0.3	+1.0(e)	-0.7



- (a) Persons missing from completed Census forms and persons in dwellings missed by Census (i.e. dwellings not known to Census) or dwellings mistakenly deemed unoccupied.
- (b) Persons included on Census forms multiple times or included in error.
- (c) A negative value indicates a net overcount.
- (d) Net overcount for the non-contact sector represents over-imputation. This column includes a small contribution from late returns and Census records with insufficient personal identifier information.
- (e) A positive difference as the net overcount in the non-contact sector was higher in 2016 compared with 2011.
- (f) Percentages are calculated using the total PES population estimate as the denominator.

The large positive difference in gross overcount in the contact sector and in net overcount in the non-contact sector in 2016, compared with 2011 (changes of 90.3% and 77.2%, respectively), are driving the lower total net undercount at the Australia level. Specifically, the larger changes in overcount (including over-imputation in the Census) have offset, to a large degree, the gross undercount in the contact sector. The positive difference for the latter changed by much less in 2016 (a change of 30.0%), compared with the overcount, thus driving the total net undercount down.

Note that net undercount for the contact sector at the Australia level is equal to gross undercount minus gross overcount. This is not true for estimates below the national level, where differences in response categories between the Census and PES, and not stated values in the Census, are also taken into account.

For a detailed discussion on the sub-components that comprise both the contact and the non-contact sectors, see **Contact Sector** and **Non-contact Sector** on the Summary tab.

## Contact Sector

### CONTACT SECTOR

Net undercount in the contact sector for a given category of person can be disaggregated into four sub-components:

- Gross undercount
- Gross overcount
- Net difference in classification
- Persons with a Census not stated category.

The following table illustrates this disaggregation for broad Country of birth categories. Descriptions of the four sub-components are provided below.

**Undercount in the Contact Sector, Country of Birth, 2016**

	PES population estimate	Census count	Gross undercount	Gross overcount	Net difference in classification	Persons with Census category not stated	Net undercount in contact sector (a)(b)
	no.	no.	no.	no.	no.	no.	no.
Born in Australia	16 124 161	15 189 883	767 068	172 412	47 422	292 201	934 278
Born overseas	6 369 428	6 017 963	383 521	102 261	-47 422	117 627	351 465
Not stated	–	409 828	–	–	–	–	–
<b>Total Persons</b>	<b>22 493 589</b>	<b>21 617 674</b>	<b>1 150 588</b>	<b>274 673</b>	<b>–</b>	<b>–</b>	<b>875 915</b>

– Nil or rounded to zero (including null cells).

(a) Components of net undercount for Country of birth do not add to the total, as those with a not stated value do not contribute to the Census counts used to calculate net undercount for a given category.

(b) The net undercount is equal to the sum of the Gross undercount, the Net difference in classification and the Persons with a Census not stated category, minus the Gross overcount.

### GROSS UNDERCOUNT

Gross undercount for a category is an estimate of the number of people who should have been counted in that category in the Census (according to PES) but were missed because:

- they were not included on a completed Census form at their usual residence or another dwelling
- their dwelling was not known to Census (e.g. it was not included on the ABS Address Register or was not located by a Census Field Officer)
- their dwelling was mistakenly deemed unoccupied on Census night.

The PES estimated that 1,150,588 persons were missed in the Census in 2016.

As Census was unaware of these persons and dwellings and therefore no imputation was performed, this component is separate from the persons imputed into non-responding dwellings, which forms part of the non-contact sector.

### GROSS OVERCOUNT

Gross overcount for a category is an estimate of the excess count of persons in that category (according to PES) either from being counted on more than one Census form or from being counted in error (e.g. where a person was overseas on Census night and should not have been counted, but was included on a Census form at their usual residence). In 2016, 274,673 persons were estimated as having been overcounted in the 2016 Census, compared with 144,341 persons in 2011.

## NET DIFFERENCE IN CLASSIFICATION

Occasionally, responses obtained for a person in the PES interview are not consistent with those obtained for the same questions in the Census. In addition, a value that was imputed in the Census for a missing response (e.g. Age, Sex, or State/Territory of usual residence) is likely to differ from the PES response.

The net difference in classification for persons actually counted in the Census contact sector is equal to the estimated number of people in the given category as reported in PES, minus the number counted in that same category in the Census. In 2016, for example, a net difference of 47,422 persons contributed to the net undercount estimates for Australia-born (according to their response in PES), even though they were counted as born overseas in the Census. For more information, see **Technical Note 1: Differences in Classification between the PES and Census** on the Explanatory Notes tab.

## PERSONS WITH CENSUS CATEGORY NOT STATED

The Census contact sector contains some dwellings which were responding in the Census but returned only a partially completed Census form. For Census purposes, values for Age, Sex, Registered marital status, and State/Territory of usual residence are imputed during Census processing in cases where these items have been left blank. Missing values for other items remain not stated in the final version of Census counts (including Indigenous status and Country of birth).

To create benchmarks for Indigenous status and Country of birth for PES estimation, persons in the contact sector with a Census category value of not stated for these items had their Census response imputed during PES processing. For details on this process, see 'Resolution of Census Not Stated Values for PES' in **Census Counts** on the Explanatory Notes tab. In 2016, there were 409,828 persons who had a missing value for Country of birth in the Census contact sector. An estimated 292,201 of these were imputed as born in Australia and therefore contributed to the net undercount for that category.

## CALCULATING NET UNDERCOUNT FOR THE CONTACT SECTOR

The net undercount for a category of person in the contact sector (e.g. those born in Australia) is equal to the sum of persons undercounted in the Census (i.e. gross undercount), persons with a net difference in classification between PES and Census, and those with a not stated response in the Census who should have been counted in the particular category, then subtracting those persons who were overcounted in the Census (i.e. gross overcount) in that category.

## Non-contact Sector

### NON-CONTACT SECTOR

The Census count in the non-contact sector comprises all person records of the following types:

- imputed persons for non-responding dwellings
- late returns
- records with insufficient personal identifier information on their Census form.

Descriptions of the three sub-components are provided below.

### Census Counts, PES Population Estimate and Net Undercount for the Non-contact Sector, 2016

	no.	%(a)
Census Count		
Imputed persons for non-responding dwellings	1 183 519	5.1
Late returns	563 046	2.4
Person with insufficient personal identified information	33 057	0.1
Total Census count in non-contact sector	1 779 622	7.6
PES population estimate	1 130 113	–
Net undercount in non-contact sector(b)	-649 509	–

– Nil or rounded to zero (including null cells).

(a) Percentages are calculated using the total Census count (23 397 296 persons) as the denominator.

(b) A negative value indicates a net overcount.

## IMPUTED PERSONS FOR NON-RESPONDING CENSUS DWELLINGS

Non-responding dwellings in the Census are dwellings where the Census never obtained a return, and the dwelling could not be established as having been unoccupied on Census night. During Census processing, imputation was used to create a number of people and their Age, Sex, Registered marital status and State/Territory of usual residence into these dwellings. Values of all other variables for these imputed people were set to not stated or not applicable, depending on the imputed value for Age. Missing values for Indigenous status and Country of birth were then imputed during PES processing. Inevitably, the imputed values differ from the true but unknown values to some extent.

The number of person records imputed into Census non-responding dwellings was higher (5.1% of all Census person records) in 2016 compared with 2011 (3.6%). Imputed person records made up the majority of the Census non-contact sector in both 2016 and 2011 (67% and 87%, respectively), although the change in 2016 is due to the higher number of late returns and the addition of Census person records with insufficient personal identifier information. Imputed person records are also the primary driver for the net overcount in the non-contact sector, which is essentially over-imputation in the Census.

## LATE RETURNS

Contact from the ABS following selection in the PES may act as a reminder to return a completed Census form for those people who have not already done so. These late returns, if not identified, would result in the PES sample having a higher proportion of Census response than in the overall population, resulting in a downward bias in the estimate of net undercount. To protect against this, and maintain the statistical independence of the PES from the Census, all Census forms received after the start of 2016 PES enumeration were deemed late.

For the purpose of PES estimation, all dwellings from which late Census forms were received were treated as though they had not been contacted in the Census and classified to the Census non-contact sector.

The number of Census late returns was higher in 2016 (2.4% of all Census person records) compared with 2011 (0.5%). This comes as a result of changes to the Census enumeration model in 2016, particularly the option for respondents to post back their Census paper form up to several weeks after the end of Census field operations, which was not available in 2011. For further information on changes to the 2016 Census, see **Major Changes to the 2016 Census** on the Explanatory Notes tab.

## CENSUS PERSONS WITH INSUFFICIENT PERSONAL IDENTIFIER INFORMATION

A key requirement for successfully linking a PES person to their corresponding Census record is a sufficient level of data quality on both sides, including minimal-to-no missing data for responding persons. Without this, the ability to link to Census (i.e. confirming a PES person was counted) is substantially reduced.

As part of the continuous improvements to PES methodology, an adjustment was introduced into the estimation method in 2016. This method identified Census records that had insufficient personal identifier information, required for linking to PES (e.g. where Census data was missing or imputed for multiple linking variables). To remove the potential for any upward bias on the PES population estimates (and level of net undercount) that would occur if these records were retained for estimation, they were moved to the Census non-contact sector and treated in a similar fashion to late returns.

The adjustment moved 33,057 Census persons to the non-contact sector, or 0.1% of all persons counted in the Census. Analysis of the impact of this adjustment showed that it reduced the estimated net undercount rate for Australia by 0.10 percentage points. This is an indicator of the extent of the upward bias on the estimated net undercount that would have existed had these records not been identified and treated appropriately.

## CALCULATING NET UNDERCOUNT FOR THE NON-CONTACT SECTOR

An estimate of net undercount in the non-contact sector for a category of person is equal to the difference between the PES estimate of the number of people who **should** have been counted in the non-contact sector in that category (as reported in the PES) and the Census count of people in that category in the non-contact sector.

## CALCULATING THE CENSUS IMPUTATION ERROR

Imputed records are the primary driver for the net overcount in the non-contact sector, and will therefore impact the overall net undercount figure for Australia. For example, if the Census count increases due to a higher number of imputed persons but all else remains the same, net undercount will decrease (and vice versa). It is therefore important that PES corrects for the Census imputation error in the non-contact sector, and that the number is reported so it can be used to better understand estimates of net undercount.

An estimate of the number of people who **should** have been imputed in the Census can be calculated by subtracting the PES estimate of late returns and for those person records with insufficient personal identifier information from the PES estimate of people who should have been counted in the non-contact sector.

The **Census imputation error** is then the difference between this PES estimate and the actual Census count of imputed people. For 2016, the imputation error is estimated to be 650,856 persons that were over-imputed by Census during data processing. This over-imputation was likely the result of some non-responding dwellings being deemed occupied when they were in fact unoccupied, as well as too many people being imputed into non-responding occupied dwellings.

## About this Release

The Census Post Enumeration Survey (PES) is a household survey conducted by the Australian Bureau of Statistics (ABS) shortly after each Census, in order to provide an independent measure of Census coverage. This release presents estimates of undercount and overcount for the 2016 Census of Population and Housing, resulting from the 2016 PES. It includes data cubes and explanatory material as well as providing net undercount estimates by a range of demographic characteristics. This publication also details how the 2016 PES was conducted and how 2016 PES estimates have been calculated.

# History of Changes

This document was added or updated on 23/02/2018.

23/02/2018

An issue has been identified with how the components of net undercount were created for the 2016 PES estimates. The contact sector should include PES respondents who were missed by Census because their dwelling was mistakenly deemed unoccupied on Census night. It was identified, however, that these respondents were erroneously treated as part of the non-contact sector. This error affected approximately 700 (unweighted) records, which have now been corrected.

Changes have been made to the contact and non-contact sector estimates. In particular,

- Net undercount in the contact sector increased by 159,334 persons, driven by an increase in gross undercount (131,813 more persons) and a small decrease in gross overcount (27,521 fewer persons)
- Net overcount in the non-contact sector increased by 159,334 persons
- The 2016 PES population estimates and total net undercount rates have not changed.

These changes can be found in the following sections of this publication:

- Key Findings
- Components of Net Undercount
- Contact Sector
- Non-contact Sector
- Table 5 in the Estimates of Net Undercount 2016 data cube

## Explanatory Notes

### The Importance of Effective Statistical Independence of PES

#### THE IMPORTANCE OF EFFECTIVE STATISTICAL INDEPENDENCE

**1** The PES is designed to be a robust and independent measure of Census coverage. Therefore, it is critical that the statistical independence between the PES and the Census is effectively managed.

**2** There are two aspects to statistical independence, both of which were effectively managed throughout the 2016 PES cycle: population independence and operational independence.

**3 Population independence** refers to the principle that there should be no subgroups of the population where being missed in the Census indicates that a person or dwelling is also more likely to be missed by the PES. In practice, some population subgroups may be less likely to respond to both the Census and the PES. Although the PES estimation process adjusts for this to some extent, by subdividing the population into smaller groups where the assumption of population independence is more likely to be true, population independence is always more difficult to achieve than operational independence.

**4 Operational independence** requires that Census operations do not influence the PES in any way, and vice versa. The operational independence of the PES from the Census was effectively monitored at every stage of the 2016 cycle, including enumeration, processing and administration. Steps taken to ensure this independence included:

- selecting the PES sample from an independent sample frame
  - using separate office staff in the PES and Census
  - ensuring PES interviewers were not employed as Census Field Officers in the same area
  - maintaining the confidentiality of the PES sample so Census field and office staff were unaware of which areas were selected in the PES
  - using a separate and secure information technology environment for processing the PES data
  - ensuring Census forms received after PES enumeration commenced (i.e. late returns) were excluded from PES estimation.
- For more information, see **Components of Net Undercount** on the Summary tab.

## Scope and Coverage of the 2016 PES

#### SCOPE AND COVERAGE OF THE 2016 PES

##### SCOPE OF THE 2016 PES

**1** For Census and PES purposes, scope refers to the group of people about which information is required. A set of rules is applied to determine whether a selected person is within this population of interest.

**2** The scope of the Census is every person present in Australia on Census night with the exception of foreign diplomats and their families. Ideally the PES would sample from all people who were, or should have been, counted in the Census. For practical reasons, however, there were a number of areas, dwellings and people outside the scope of the 2016 PES.

**3** Of the people present in Australia at the time of the PES, the following are not included in the scope of the PES:

- overseas visitors who were not in Australia on 9 August 2016 (Census night)
- foreign diplomats and their families
- people in non-private dwellings such as hotels, motels, hospitals and other institutions
- homeless people (as the sample selected in the PES is based on a selection of dwellings)
- babies born after 9 August 2016
- people in Other territories (e.g. Christmas Island, Norfolk Island and Jervis Bay Territory) or Australian external territories (e.g. Australian Antarctic Territory, Coral Sea Islands Territory).

**4** The PES also does not obtain information about people who died between Census and PES.

#### **Remote Areas and Discrete Communities**

**5** The 2016 PES sample included remote areas and Discrete Communities, which ensured a more complete geographic coverage of the PES. In 2016, the risk to statistical independence in these areas and communities was effectively managed through careful monitoring of Census field operations, interviewer training and well established field procedures. No PES interviewers worked on Census collection activities in these same areas.

#### **Non-private Dwellings (NPDs)**

**6** The 2016 PES sample excluded people living in NPDs, as has been the case in previous PES cycles. NPDs are establishments which provide predominantly short-term accommodation for communal or group living, and often provide common eating facilities. They include hotels, motels, hostels, hospitals, religious institutions providing accommodation, educational institutions providing accommodation, prisons, boarding houses and short-stay caravan parks.

**7** The inclusion of NPDs was considered for the 2016 PES; however, they were excluded for the following reasons:

- a higher number of NPDs are found to be vacant at any given point in time, compared with private dwellings, due to their short-stay nature
- many of the people who are contacted in these dwellings are not in scope of the PES; e.g. overseas visitors who were not in Australia on 9 August
- many of the people usually live in a private dwelling and could be selected at their usual residence
- the quality of information is often poor, especially in institutions where information cannot be provided by the respondents themselves and is collected from staff or administrative lists
- there is an increased risk to the statistical independence of the PES for institutions, as Census uses the same staff or administrative lists
- the costs of enumerating them is prohibitive compared with private dwellings due to the additional time and effort to enumerate.

#### **COVERAGE OF THE 2016 PES**

**8** Coverage refers to a set of rules designed to give each in scope person in the PES sample a single chance of selection in the survey. These rules are implemented by associating each person with a single dwelling through a series of questions in the PES interview, such as where each person usually lives and whether they (or anyone else) are staying at their usual residence during the PES enumeration period.

**9** An example of the need for coverage rules is for a visitor staying at a dwelling selected in the PES. If the visitor reports that someone else is staying at their usual residence during PES enumeration, or they return to their residence at a later point during the enumeration period, this visitor could have two chances of being included in the PES (i.e. once at their usual residence and again at their current location). In this example, the PES coverage rules would associate the visitor with their place of usual residence and not the dwelling they were visiting, thereby giving them only one chance of being included in the PES enumeration.

## **Major Changes to the 2016 PES**

#### **MAJOR CHANGES TO THE 2016 PES**

**1** There were a number of changes implemented for the 2016 PES, some of which resulted from the changes made to the 2016 Census.

**2** Key changes made to the PES field collection phase were:

- A sample size increase of 20%. This was done to ensure the high level of accuracy for the PES estimates (as measured by the sampling error) was maintained, as well as to manage the risk of a number of operational unknowns when the PES sample was being designed (e.g. potential overlap in Census and PES enumeration) that could have significantly impacted the PES response rates and sample size
- The focused sampling strategy to improve estimates for the Aboriginal and Torres Strait Islander population was removed, as a review in 2015 indicated that most of the improvement in the estimates for this population was achieved through other means
- Follow-up phases for the 2016 Census started and finished several weeks later than the 2011 Census. Accordingly, the 2016 PES enumeration started later than in 2011 and commenced immediately following conclusion of the Census follow-up phase

- The PES enumeration period was extended to cater for the larger sample and the increasing difficulty in making contact with householders
- A telephone interview option was introduced to provide more convenience and choice for PES respondents. Respondents were able to register for a telephone interview via a number of methods (online or via telephone) by following instructions provided in the Primary Approach Letter. The introduction of telephone interviewing improved the effectiveness of enumerating the large 2016 PES sample in a short timeframe.

**3** Key changes made in the PES processing and estimation phase include:

- Address text matching was introduced to provide an opportunity to link PES dwellings to a specific entry on the ABS Address Register
- For the purposes of PES estimation, the respondent's age (which is collected as at the time of the PES interview) was backdated to 9 August 2016 (i.e. their age on Census night) using date of birth, where provided
- Census records that had insufficient personal identifier information, thus reducing the likelihood of linking to PES, were identified and moved to the Census non-contact sector for the purposes of PES estimation.

## **PES Collection Methodology**

### **PES COLLECTION METHODOLOGY**

#### **FIELD PROCEDURES**

**1** The 2016 PES required careful planning and preparation, due to the extended collection period of the 2016 Census, to ensure that data quality and response rates were not compromised.

**2** Various field strategies were devised for the enumeration of the PES. Standard procedures were used, where possible, when enumerating private dwellings and in Discrete Communities. In some cases, however, it was necessary to modify these procedures. This was particularly true in Discrete Communities, where changes were made to account for language and cultural issues while still ensuring the underlying concepts remained the same across both sample components.

#### **PES FIELD PROCEDURES TESTING**

**3** The PES adapted a standard testing regime, with all field tests undertaken as part of Census test activities. The PES tested new and existing collection methodology and field procedures via the following activities:

- 2014 Major Test – tested the viability of an online Household Contact Details Form in preparation for telephone interviewing
- 2015 Field Test – tested the questionnaire for both the General Population and Discrete Community samples, and tested telephone interviewing for the General Population
- 2016 respondent material testing – tested the suitability and functionality of all respondent facing materials (Primary Approach Letter and Flyer) with internal and external specialists.

**4** Efficiencies identified as a result of the various tests were implemented before the 2016 PES enumeration commenced.

### **PES ENUMERATION**

#### **General Population Sample**

**5** Specially trained PES interviewers collected data through interviews that started seven weeks after Census night. Interviews were conducted with any responsible adult of the household, who was asked to respond on behalf of all household members.

**6** Most interviews were conducted face-to-face; however, respondents were given the option to register for a telephone interview, which had a take-up rate of 32%. Respondents were provided registration instructions on a Primary Approach Letter delivered by mail to the selected dwelling approximately one week prior to the commencement of interviewing.

**7** The PES collection methodology differed from the way Census collected its information, where most forms were self-completed. A major advantage of interviewer-administered questionnaires is that people can be provided with assistance if they are uncertain about the meaning of questions, and help is also given to ensure no questions are left unanswered.

**8** All dwellings were enumerated using Computer Assisted Interviewing (CAI).

#### **Discrete Community Sample**

**9** Enumeration of the Discrete Community sample was staggered over time and undertaken as soon as practicable after the completion of Census collection activities in that community. Interviews in Discrete Communities were conducted by specially trained PES interviewers with the assistance of facilitators recruited from within the selected community. The facilitator assisted in establishing rapport with respondents, helped identify residents of the selected households, and provided interpretation when needed.

**10** To preserve the independence of the Census and the PES, every effort was made to recruit facilitators who were not involved in Census collection. Where this was not possible, PES interviewers ensured facilitators had only a limited role in the PES interview and provided assistance only where necessary, such as with language interpretation. In some communities it was acceptable to enumerate without the help of a facilitator, but generally they played an important role in being able to effectively enumerate in the communities.

11 All dwellings were enumerated with a customised questionnaire using CAI.

## **CENSUS FOLLOW-UP**

### **General Population Sample**

12 Intensive Census follow-up procedures were employed at the end of the Census collection period from dwellings that did not return a Census form within the required time frame.

13 All Census follow-up activities for the General Population were closely monitored by PES staff to ensure the risk of overlap between the Census and PES, and hence the statistical independence between the two, was effectively managed. All follow-up was completed by Friday 23 September 2016. The first Primary Approach Letters (PAL) for PES were received by respondents on Monday 19 September. PES field enumeration and interviewing then commenced on Sunday 25 September 2016 (i.e. after the completion of Census follow-up) and continued until Wednesday 9 November 2016.

### **Discrete Community Sample**

14 PES enumeration in Discrete Communities commenced on 22 August 2016 and was completed on 11 November 2016. As PES enumeration of Discrete Communities had a staggered start date (based on when Census field operations in each community were complete), there was no overlap between PES enumeration and Census collection in the selected communities.

## **QUESTIONNAIRE**

### **General Population Sample**

15 The PES questionnaire collected personal details (name, sex, date of birth, age, relationship in household, registered marital status, country of birth and Indigenous status) to facilitate the matching of PES person records to Census person records during processing. This enables accurate undercount estimates to be generated for Age, Sex, Indigenous status, and a number of other characteristics.

16 For each person in the selected dwelling, the PES also asked:

- whether they were included on a Census form (and if so, where)
- whether they could have been included on a Census form at other addresses (and if so, where)
- where they stayed on Census night
- a set of questions to determine whether they **should** have been counted in the Census.

17 The different addresses collected in the PES were used to search Census records to determine the number of times each PES respondent was counted in the Census. Visitors to households included in the PES were also asked for their address of usual residence.

### **Discrete Community Sample**

18 As in 2011, a specially designed questionnaire was used in Discrete Community enumeration. A number of questions in the General Population PES questionnaire were not considered applicable to people living in Discrete Communities and were adapted to ensure information was collected in the most culturally appropriate manner. For example, it is more common for Aboriginal and Torres Strait Islander persons in a community to be known by more than one name. This tailored questionnaire ensured sufficient information was collected to effectively match persons, taking into account the potentially different nature of matching.

19 A Community-level questionnaire was also completed for each selected Discrete Community with the help of the Community contact or council officer. Information collected on this questionnaire, such as whether a significant event occurred at the time of the Census (e.g. a sports carnival), was used to help with respondent recall and assisted in the completion of parts of the individual questionnaires.

## **2016 PES Sample and Response Rates**

### **2016 PES SAMPLE AND RESPONSE RATES**

#### **PES SAMPLE DESIGN**

1 The 2016 PES sample comprised two components: private dwellings (houses, apartments, etc.) within the General Population sample, and private dwellings within the Discrete Community sample.

2 A total of 53,800 dwellings were randomly selected in the PES sample design. Taking into account sample loss and non-response expectations, approximately 43,000 dwellings were expected to respond to the PES, including around 800 dwellings from 33 Discrete Communities in NSW, Qld, SA, WA and NT. This represented an increase of 20% on the size of the 2011 PES sample.

#### **GENERAL POPULATION SAMPLE**

3 The General Population sample was selected through a stratified, multi-stage cluster design from the private dwelling framework of the ABS Population Survey Master Sample. Areas are grouped into strata based on locality, population density, remoteness,

projected population growth and socio-economic characteristics. Each private dwelling within the stratum has the same probability of selection.

4 A focused sampling strategy was introduced into the General Population sample in 2011 to improve estimates for the Aboriginal and Torres Strait Islander population. The strategy was reviewed in 2015, which showed that most of the improvement in Aboriginal and Torres Strait Islander estimates was achieved by increases to the overall sample size, rather than from the focused design. As such, the strategy was discontinued in 2016.

#### DISCRETE COMMUNITY SAMPLE

5 For the purposes of PES sampling, Discrete Communities were grouped into sets comprising main communities and their associated outstations. The selection of main communities was undertaken with probability proportional to the size of the set. The aim was to select as representative a sample as possible while also considering cost constraints, reasonable interviewer workloads and expected sample size.

6 The sample selection process within the Discrete Community required the interviewer to compile a list of all the dwellings within the community. From this, a sample was taken using systematic equal probability sampling.

7 A selection of outstations associated with each selected main community was also included in the sample. Each outstation had an equal chance of selection and, once selected, all dwellings within the outstation were enumerated. Occasionally, an outstation was not enumerated if it was difficult to access from the main community resulting in a small loss of coverage.

#### 2016 RESPONSE RATES

8 The total number of fully responding dwellings across both sample components, after allowing for sample loss, was 42,463. This represented a response rate of 91.2% for the General Population sample and 92.7% for the Discrete Community sample. This is a lower than 2011, which achieved 94.0% and 96.4%, respectively.

#### APPROACHED SAMPLE, FINAL SAMPLE AND RESPONSE RATES

	NSW	Vic.	Qld	SA	WA	Tas.	NT	ACT	Aust.
<b>GENERAL POPULATION</b>									
Households approached (after sample loss)	10 267	8 802	8 263	4 920	5 587	2 779	3 308	1 698	45 624
Households in sample	9 147	8 034	7 634	4 587	5 211	2 638	2 791	1 555	41 597
Response rate (%)	89.1	91.3	92.4	93.2	93.3	94.9	84.4	91.6	91.2
Persons in sample	24 713	21 153	20 434	11 519	13 677	6 230	7 921	4 177	109 824
<b>DISCRETE COMMUNITY</b>									
Households approached (after sample loss)	28	–	124	65	153	–	564	–	934
Households in sample	22	–	115	65	137	–	527	–	866
Response rate (%)	78.6	–	92.7	100.0	89.5	–	93.4	–	92.7
Persons in sample	52	–	465	313	578	–	2 811	–	4 219

– Nil or rounded to zero (including null cells).

## Linking and Matching



## LINKING AND MATCHING

1 A linking and matching exercise is used to determine whether each PES respondent was counted in the Census (and how many times), whether they were counted in error, or whether they were missed entirely. Linking PES persons to their Census form involved a range of automated and manual processes, focused on finding matches between approximately 114,000 PES person records and over 22 million Census person records (not including records for the imputed persons in non-responding dwellings).

2 The various processes that made up linking and matching in 2016 were:

- Standardisation
- Address coding
- Address text matching
- Automated Data Linking (ADL)
- Clerical review using the Match and Search System (MSS).

## STANDARDISATION

3 In preparation for ADL, PES data were repaired and standardised through a four-stage process to convert them into a format that could be directly compared with similarly standardised Census data:

- **Data Repair** – cleaned the data by removing non-alphabetic characters and capitalising the remainder, and by removing additional spaces
- **Name standardisation** – converted common nicknames, abbreviations, misspellings or variations on a name to their 'origin name' (e.g. Beth, Eliza and Libby were converted to Elizabeth)
- **Address text parsing** – cleaned raw address information by cross-referencing the components of an address, such as street name or suburb, with known street names or suburbs
- **Data transformation** – ensured that each variable was comparable to its Census counterpart (e.g. ensuring PES numeric identifiers for Indigenous status matched to those of Census).

## ADDRESS CODING

4 Address information is essential for matching PES respondents to their Census form. PES addresses were divided into two categories:

- **Enumeration Address** – the address at which the PES interview took place
- **Search Address** – all other addresses, including the usual address of visitors to the PES dwelling, the address at which the PES person was located on Census night, the address at which they were included on a Census form, and any other addresses where the respondent may have been included on a Census form.

5 The AddressCoder@ABS is a web-based application used to geocode each type of address. From this, geographic information was assigned, such as a Census Field Area (CFA), a Mesh Block (MB), or a Statistical Area 1 (SA1) boundary, which were all used during automated data linking of persons.

6 Geocoding via the address coder was relatively resilient to errors in the address text (e.g. character substitution or form-scanning errors) as it needed only to identify the locality and not a specific address or dwelling.

7 Addresses that were unable to be coded automatically via the AddressCoder@ABS application were sent to a processing team for manual coding. This manual process utilised various methods, including mapping software, to thoroughly scrutinise addresses and achieve the most accurate geographic coding possible.

## ADDRESS TEXT MATCHING

8 Address text matching was introduced in the 2016 PES and provided an opportunity to identify potential dwelling links based on exact address information. It was used to match a PES address to a specific entry on the ABS Address Register.

9 This exercise was particularly useful for dwelling types that were in scope for linking (e.g. unoccupied dwellings) but unable to be linked via automated data linking, which is person-based. The proposed dwelling link was then fed through to the clerical matching process for confirmation. It should be noted, however, that address text matching was susceptible to errors or missing entries in the address register.

## AUTOMATED DATA LINKING (ADL)

### Linking

10 ADL refers to the use of probabilistic linking methods to determine possible links between Census and PES data in an automated fashion, before any clerical matching process begins. This was introduced as the primary linking method in the 2011 PES, which used the Freely Extensible Biomedical Record Linking (FEBRL) software, and was used again in 2016.

11 The automated linking process used a range of personal and address characteristics to evaluate the chance that a PES record and a Census record were for the same person. The method generated large numbers of candidate links and then used a process of elimination to filter down to genuine matches.

12 Seven different linking runs were used in ADL to compare PES and Census records, each focused on a slightly different combination of name, addresses and demographic variables. At the beginning of each run, a list of PES and Census records was

obtained by selecting a subset of the PES and Census datasets which agreed on a small number of variables (e.g. the same SA1, date of birth, and surname). This process, called 'blocking', reduced the number of Census and PES records to compare within a run, in order to increase the likelihood of proposing good quality links.

**13** The 2016 PES used a set of blocking variables that were comparable to those used in 2011, allowing for updated geographical classifications. The seven linking runs used various combinations of the following:

- SA1
- CFA
- State
- Postcode
- Standardised name (blocking on initials)
- Birth date (day, month and year)
- Sex
- Registered marital status.

**14** Potential links were generated by assigning weights to reflect the level of agreement for combinations of linking variables within each block. Large positive weights indicated probable matches, while large negative weights indicated probable non-matches.

### **Consolidation of ADL links**

**15** A series of processes was undertaken following the ADL runs to clean and consolidate the proposed links.

**16** The Collect, Analyse, Reduce, De-duplicate and Systematise (CARDS) process identified and rated the most plausible links from each ADL run for all PES respondents. The process then combined the links from all ADL runs and removed any duplicates, with links from earlier runs taking precedence.

**17** The final step of the automated linking process was to group person links together into dwelling units when they were co-located in the same PES-Census dwelling pair, through a process called Dwelling Link Rating (DLR). This had several benefits including:

- finding additional candidate person links by upgrading lower quality links if they were co-located with high quality person links
- grouping links by whole dwellings in preparation for clerical review. The MSS was more effective when handling person links in the same dwelling, as the processor could examine the entire household and their corresponding Census form at the same time.

**18** The proposed dwelling links were then categorised into the following:

- **Platinum** – the dwelling link was of sufficiently high quality that it could be immediately confirmed and did not require clerical review
- **Silver** – the dwelling link was of moderate quality (e.g. there were high quality links for some persons in a dwelling and not others) and required clerical review to confirm or reject the quality of the link
- **Tin** – the link was of low quality and was unlikely to be a true match, therefore clerical review was required to search for matches for these dwellings and persons without the assistance of ADL.

**19** All PES dwellings with either Silver or Tin links were sent for clerical review. A small percentage of Platinum links were also clerically reviewed for quality assurance purposes.

### **MATCH AND SEARCH SYSTEM (MSS)**

#### **Processing in the MSS**

**20** While ADL is a critical component of PES linking and matching, it cannot entirely replace the traditional clerical decision-making process. Clerical judgment will always be required to resolve the more complex or ambiguous cases and provides a means of quality assuring the automated processes. The MSS is used for this purpose.

**21** The MSS allows processing staff to manually search, view, and compare PES and Census data. There are two phases of processing in the MSS:

- Evaluate a candidate link provided by ADL and confirm or reject the link for both dwellings and people between PES and Census
- A clerical search for a link in the absence of a good ADL candidate link, by searching for people on Census forms at alternative search addresses provided by the PES respondent.

**22** To evaluate ADL links, the processor first needed to confirm whether the ADL-proposed dwelling link was correct. Once the dwelling link was confirmed, the Census person records for that dwelling were compared with the PES person records, using information such as Name, Sex, Date of birth, Age, Registered marital status, Indigenous status and Country of birth. The extent to which each of these variables was the same, in both the PES and the Census, determined a match or a non-match status for the pair.

**23** Where the ADL-proposed dwelling link was rejected, or if no dwelling link was proposed by ADL, processing staff undertook an intensive search. This search focused on the nominated (and surrounding) CFAs for all search addresses provided by

respondents during the PES interview, in order to locate possible Census forms where that person was included. If a dwelling match was found, they proceeded to rate the candidate person matches within that dwelling as per the above.

**24** Some redevelopment of the MSS was necessary in 2016 to ensure the system aligned with the changes made to the 2016 Census enumeration model. During this redevelopment work, some system enhancements were made to further strengthen and streamline the clerical matching processes and outcomes; however, the system is considered to be largely comparable with the 2011 version.

### MSS Quality Assurance and Adjudication Processes

**25** Quality assurance (QA) procedures were used to ensure the accuracy of MSS outcomes. For example, all records sent to the MSS were processed twice. The QA workloads were processed by a different processor, and there were no identifiers to mark it as an original or QA workload.

**26** Where the original and the QA match status corresponded, the original match status was accepted. Where there was a discrepancy between the original match status and the QA match status (at either the dwelling or person level), the records were flagged for adjudication by a senior processor (adjudicator) who reviewed all information and determined which match status was correct. Where both the original and QA records were deemed to be incorrect, the adjudicator reprocessed the record.

**27** The adjudication process was also useful in identifying potential issues or areas where processing staff were having difficulty. This allowed ongoing feedback to be provided to the MSS staff and contributed to the overall quality of PES processing.

**28** A 5% sub-sample of Platinum ADL linked dwellings and persons was also processed fully through the MSS as a quality measure of the ADL. Reprocessing these records confirmed the robustness and high quality of the ADL links. Interrogation of any high quality ADL links that were rejected by the processing staff was also undertaken as a further quality assurance measure.

### DISCRETE COMMUNITY PROCESSING

**29** As per 2011, ADL was not utilised for processing the Discrete Communities sample in 2016. The low quality geocoded data for these areas, as well as the ability for PES respondents to provide alternative names, would have complicated the ADL process. Instead, the sample underwent full clerical searching and matching in the MSS. System enhancements made to the MSS enabled more thorough and efficient processing to be completed.

**30** The process involved searching the entire community for a person match, rather than just searching within a single dwelling. Person matching in Discrete Communities used the same rules for determining a match as in the General Population, but allowed for the use of up to two alternative names for each person when matching on name.

### CONFIDENCE OF MATCH DECISIONS

**31** Outcomes from linking and matching processes underwent a high level of scrutiny and quality assurance in 2016, to ensure the PES did not miss links for PES persons who were actually counted in the Census, and did not link a PES person to a Census record in error.

**32** Final match rates for the General Population for persons with at least one link to 2016 Census were lower than the 2011 equivalents (91.7% and 92.7%, respectively). This change was driven by a reduction in Census response rates. However, more high quality links were found by ADL in 2016 that did not require clerical review, compared with 2011 (65.1% and 59.8%, respectively). This is likely to be the result of improved capture of text fields, such as names, from increased online Census uptake.

### Matching Outcomes, 2011- 2016

		2011		2016	
		no.	%	no.	%
<b>GENERAL POPULATION</b>					
Persons matched(a)		87 645	92.7	100 783	91.7
ADL Platinum (not clerically reviewed)		52 398	59.8	65 576	65.1
ADL Silver (clerically reviewed)		34 653	39.5	33 776	33.5
Intensive search		594	0.7	1 431	1.4
Persons not matched		6 894	7.3	9 142	8.3
<b>Total Persons</b>		<b>94 539</b>	<b>100.0</b>	<b>109 925</b>	<b>100.0</b>
<b>DISCRETE COMMUNITY</b>					
Persons matched(b)		2 528	87.1	3 394	79.2
Persons not matched		373	12.9	892	20.8
<b>Total Persons</b>		<b>2 901</b>	<b>100.0</b>	<b>4 286</b>	<b>100.0</b>

(a) A person can be matched more than once, therefore the total number of matches does not sum to the total number of matched persons.

(b) Matches for the Discrete Community sample were made via the MSS only.

## Estimation

### ESTIMATION

**1** PES estimation involves assigning a weight to each selected PES dwelling and then to each person for whom a PES response was obtained. The weights attached to PES persons allow the PES sample to represent the whole population of interest; i.e. all usual residents in Australia on Census night, including people in non-private dwellings (e.g. hotels, hospitals and jails) which are

not covered by the PES dwelling sample.

## DWELLING WEIGHTING

2 Dwelling weighting for the 2016 PES comprised two stages. For private dwellings selected in the PES that were found in the Census, the first stage of weighting adjusted the PES selection weight (the inverse of the probability of a dwelling being selected in the PES sample) such that the adjusted weights added up to the Census private dwelling count within categories based on geography and dwelling characteristics. A first-stage weight adjustment was also applied to private dwellings selected in PES that were missed in the Census. For dwellings in the Discrete Community sample, a similar first-stage weight adjustment was applied based on dwelling counts for communities within each state and territory.

3 The second stage of dwelling weighting applied a non-response adjustment so that the responding PES dwellings represented other dwellings from which no response was obtained.

## PERSON WEIGHTING

4 The initial stage of person weighting adjusted the dwelling weights to ensure that the PES estimates of people counted (in the Census) in private dwellings and Discrete Community dwellings (other than the non-contact sector) in a set of benchmark categories matched the actual Census counts for these categories. The weight adjustment applied to a person did not depend on whether they responded in the Census, but only on characteristics of the person as reported in the PES.

5 As a final step in weight adjustment, the initial person weights were adjusted so that the PES estimates also represented people in non-private dwellings, such as hotels, hospitals and jails, which were not covered by the PES. The person weighting step in PES processing calculates weights for all PES records, including those relating to the non-contact sector.

6 Intuitively, a good set of weights for the PES should ensure that if the PES were used to estimate the actual Census count, the PES would get the right answer. The above step ensured this was the case. Technically, this is a desirable property for a set of PES weights to have since there is a very strong relationship between the actual Census count and the count that the Census should have made.

7 In 2016, the ABS again used the Prediction Regression (PREG) estimator, which was developed and used in 2006 and 2011. This method was introduced in 2006 to account for overlapping benchmark categories and the situation where people gave different responses between PES and Census. The resulting estimates are categorised by the PES response (i.e. where they should have been counted), rather than the Census response. A detailed description of the PREG estimator can be found in Chipperfield et al. (2016)[1].

## ESTIMATES OF THE POPULATION AND NET UNDERCOUNT

8 The PES population estimate for a category of person (e.g. males) is equal to the sum of person-weights of those persons who should have been counted in that category in the Census. The final PES estimate is then adjusted for consistency with Census counts, by adding on the persons who **were** counted in that category in responding Census dwellings (i.e. dwellings from the contact sector) and subtracting the weighted estimate of these people.

9 Net undercount for any category of person is then the difference between the final PES population estimate for that category and the actual Census count (including imputed persons in non-responding dwellings). This calculation takes into account the components described in **Components of Net Undercount** on the Summary tab.

## ENDNOTES

1 Chipperfield J, Brown J and Bell P 2016. 'Estimating the Count Error in the Australian Census', Journal of Official Statistics, vol. 33, pp. 1–17.

# Scope of the 2016 Census

## SCOPE OF THE 2016 CENSUS

1 The Census aims to count every person who spent Census night in Australia, with the exception of foreign diplomats and their families. Visitors to Australia are in scope regardless of how long they have been in the country or how long they plan to stay. Australian residents out of the country on Census Night are out of scope in the Census.

2 The scope of the Census included people in the six states, the Northern Territory, the Australian Capital Territory, Jervis Bay Territory, and the Territories of Christmas Island and Cocos (Keeling) Islands, Antarctica and Norfolk Island. The inclusion of Norfolk Island in Other Territories was new for 2016, following an amendment to the **Acts Interpretation Act, 1901**. The other Australian external territories (minor islands such as Heard Island and McDonald Island) remained outside the scope of the Census. People outside Australia who were not required to undertake migration formalities, such as those on oil and gas rigs off the Australian coast, were in scope.

3 All private dwellings, except diplomatic dwellings, were included in the Census, whether occupied or unoccupied. Caravans, cabins and tents in caravan parks, and manufactured homes in manufactured home estates were counted only if occupied. Occupied non-private dwellings, such as hospitals, prisons, hotels, etc., were also included.

# Major Changes to the 2016 Census

## MAJOR CHANGES TO THE 2016 CENSUS

**1** The 2016 Census introduced significant changes to the way the Census was conducted, moving to a 'Digital First' approach. The new approach changed the way Census materials were delivered and information returned by the public.

**2** In most areas of Australia, households were mailed information in advance, which included a unique login number and instructions on how to complete the Census online. People in these households were able to request paper forms if they did not wish to complete the Census online. In other areas, where a greater need for paper forms was anticipated, these were delivered together with a prepaid mail-back envelope as the first contact, together with a login number and instructions on how to complete the form online if they wished.

**3** This change from the traditional drop off and pick up of paper forms by Census Field Officers for the majority of Australia was the most significant change in collection methodology since the first Australian Census in 1911. Central to the new approach was the development of the ABS Address Register as a frame for this large scale mail-out of Census materials.

**4** As with previous Censuses, a range of targeted enumeration strategies were developed to optimise accessibility and inclusion for population groups at risk of being missed, or where there were known social, cultural and/or physical barriers.

**5** For further information on the 2016 Census content, collection operations, the ABS Address Register, confidentiality and privacy, and processing, see the ABS publication *Census of Population and Housing: Nature of Content, Australia, 2016* (cat. no. 2008.0), released on 20 August 2015.

## Census Counts

### CENSUS COUNTS

#### USUAL RESIDENCE

**1** The Australian Census counts people where they stayed on Census night. This means the Census was conducted on an actual location or place of enumeration basis.

**2** For usual residents of Australia, place of usual residence for the 2016 Census was defined as the address at which a person has lived, or intends to live, for six months or more in 2016. While for most people their usual residence was the same as their actual location on Census night, some people spent Census night at a place other than where they usually lived. Thus, their place of enumeration and their place of usual residence were different.

**3** People visiting Australia on Census night were included in the Census counts on a place of enumeration basis but not those on a place of usual residence basis.

**4** Usual residents of Australia who were temporarily overseas on Census night were not included in Census counts on either a place of usual residence or place of enumeration basis. Counts of these people were accounted for, however, in the Estimated Resident Population (ERP) of Australia. For information on the calculation of ERP, see the ABS publication *Australian Demographic Statistics, December 2016* (cat. no. 3101.0), released on 27 June 2017.

**5** Census counts presented in this publication are on a place of usual residence basis (unless otherwise noted).

#### IMPUTATION

**6** In addition to complete person records, Census counts include:

- persons whose Census form was partially completed
- imputed person records for non-responding dwellings.

**7** During Census processing, values for Age, Sex, Registered marital status, and State/Territory of usual residence were imputed if they were left blank on the Census form (i.e. missing item response from an otherwise responding person). These variables are needed for population estimates, so they were imputed using other information on the Census form. These were also the only variables imputed for the imputed persons in non-responding dwellings.

**8** Values for all other variables left blank (e.g. Indigenous status, Country of birth or Occupation) were set to not stated or not applicable, depending on the (possibly imputed) age of the person.

**9** While the method of imputation remained the same as 2011, the 2016 Census experienced a reduction in the amount of quantitative information they could use in their imputation model. Specifically, the 'credible source' information collected for non-responding dwellings (where possible) in previous Censuses was no longer collected. This meant that information, such as whether the dwelling was occupied on Census night and the number of males and females in that dwelling, was no longer available to guide the imputation process. Instead, imputation was based solely on the values observed for responding persons.

**10** For further information about imputed persons and how they are treated in PES, see **Components of Net Undercount** on the Summary tab.

#### RESOLUTION OF CENSUS NOT STATED VALUES FOR PES

**11** The PES uses Census data items to form benchmark categories for weighting and estimation purposes, including Indigenous status and Country of birth. In cases where one or both of these two items had been set to not stated in the Census, a value was imputed during PES processing so they could be used for benchmarks in PES estimation.

**12** The imputation method involved imputing both variables together. For benchmarking purposes, 12 non-overlapping categories were defined (e.g. Australian born and not of Aboriginal or Torres Strait Islander origin, or born in one of the ten highest ranked countries and not of Aboriginal or Torres Strait Islander origin). Imputation was performed within a number of imputation classes based on geography (SA2), the Census form type, age and sex. The imputed value was based on the proportion of stated values of respondents in the same imputation class. This method also required the assumption that anyone who was imputed as Aboriginal or Torres Strait Islander was also born in Australia.

## CENSUS DWELLING TYPES

**13** Although all persons resident in Australia should have been counted in the Census, not all dwellings would have received a Census form. This is because not all dwellings were habitable or they were missing from the ABS Address Register, or, in the case of a diplomatic dwelling, did not contain people within scope of the Census.

**14** Census defines private dwellings as structures established for self-contained accommodation. A private dwelling is most often a house or flat. It can also be a caravan, houseboat, tent, or a house attached to an office, or rooms above a shop. Private dwellings are either occupied or unoccupied:

- **Occupied** – private dwellings that were occupied by one or more persons on Census night
- **Unoccupied** – private dwellings which are habitable but were not occupied on Census night. This includes unoccupied holiday houses, vacant houses, and unoccupied apartments. Unoccupied dwellings in caravan/residential parks, camping grounds, marinas, and manufactured home estates were not counted in the Census, with the exception of unoccupied residences of managers/caretakers of such establishments.

**15** Private dwellings may also be classified as non-responding or a refusal in the Census:

- **Non-responding** – private dwellings where the Census Field Officer was unable to make contact with a householder within five visits, and was unable to verify that the dwelling was unoccupied on Census night. This excludes dwellings where the householder contacted the ABS to advise that the dwelling was unoccupied
- **Refusal** – private dwellings where the householder refused to participate in the Census.

**16** For PES purposes, dwellings that provided a Census form after the commencement of PES enumeration were flagged as late returns for PES estimation. For information on the treatment of late returns, see **Components of Net Undercount** on the Summary tab.

## CENSUS DATA QUALITY

**17** Census data are subject to a number of inaccuracies resulting from errors by respondents or mistakes in collection or processing. While many of these are corrected by careful processing procedures, some remain. The effect of the remaining errors is generally slight, although it may be more important for smaller groups in the population.

**18** The main kinds of errors occurring in the Census are:

- **Partial non-response** – When completing their Census form, some people do not answer all questions which apply to them. When this occurs, some items might be imputed while others are set to not stated or not applicable, as discussed in the section on Imputation, above
- **Processing error** – Much of the Census data is recorded using automatic processes, such as scanning, Intelligent Character Recognition and other automatic processes. Quality assurance procedures are used during Census processing to ensure processing errors are kept to an acceptable level. Sample checking is undertaken during coding operations and corrections are made where necessary
- **Respondent error** – The Census is self-enumerated, meaning householders are required to complete the Census form themselves, rather than having the help of a trained interviewer, and may be completed by one household member on behalf of others. Errors can be introduced if the respondent does not understand the question, or does not know the correct information about other household members. Self-enumeration carries the risk that wrong answers could be given, either intentionally or unintentionally.

**19** For further information on sources of error in the Census, see the ABS publication Census of Population and Housing: Census Dictionary, 2016 (cat. no. 2901.0), released on 23 August 2016.

## Technical Note 1 : Differences in Classification Between the PES and Census

### TECHNICAL NOTE 1: DIFFERENCES IN CLASSIFICATION BETWEEN THE PES AND CENSUS

**1** Occasionally, the answers obtained for a person in the PES interview are not consistent with the answers obtained for the equivalent questions in the Census. There are a number of reasons why a response may differ, including:

- a person may have difficulty answering a question for themselves or another household member, either in the Census or the PES

- a person may interpret the question differently in the Census, where forms are self-enumerated for the majority of Australia, compared with the PES, which is administered by a trained interviewer
- different people may provide Census and PES responses
- the correct response could change between the Census and PES. Changes in age are taken into account using the actual date of birth, but other changes (e.g. if the person is married or divorced), may not be identified
- the Census may contain a not stated or imputed response while the PES will have a valid response.

2 For data items collected in both Census and PES, this difference in classification is most prevalent for Indigenous status and Country of birth, as these are missing in a number of Census forms. Imputation is not carried out for either of these variables in the Census where they have been left blank or for those persons imputed into non-responding dwellings.

3 The two tables below contain unweighted PES counts of the differences in classification between PES and Census for Indigenous status and Country of birth responses. These tables are counts of all responding people in the PES who have been matched to their Census record.

4 The majority of person matches between PES respondents and their corresponding Census records show the same responses for Indigenous status. For example, 94.6% of matched people who reported as Indigenous in the PES also reported as Indigenous in the Census, and 98.8% of those who reported as non-Indigenous in the PES also reported as non-Indigenous in the Census.

**Difference In Classification(a), Indigenous status, 2016**

Census response	PES response					
	Indigenous		Non-Indigenous		Total	
	no.	%	no.	%	no.	%
Indigenous	5 741	94.6	381	0.4	6 122	6.0
Non-Indigenous	282	4.6	93 972	98.8	94 254	93.1
Not stated	45	0.7	803	0.8	848	0.8
<b>Matched Persons</b>	<b>6 068</b>	<b>100.0</b>	<b>95 156</b>	<b>100.0</b>	<b>101 224</b>	<b>100.0</b>

(a) Excludes 1,554 Indigenous persons and 8,942 non-Indigenous persons in the PES who were not matched to Census.

5 Similar results can be seen for Country of birth, where the majority of persons reported the same birth place in both PES and the Census. For example, 97.7% of matched people who reported as born overseas in the PES were counted as born overseas in the Census, and 97.7% of matched people who reported as Australian born in the PES were also counted as Australian born in the Census.

**Difference In Classification(a), Country of Birth, 2016**

Census response	PES response					
	Australia		Overseas		Total	
	no.	%	no.	%	no.	%
Australia	73 444	97.7	199	0.8	73 643	72.8
Overseas	385	0.5	25 461	97.7	25 846	25.5
Not stated	1 332	1.8	403	1.5	1 735	1.7
<b>Matched Persons</b>	<b>75 161</b>	<b>100.0</b>	<b>26 063</b>	<b>100.0</b>	<b>101 224</b>	<b>100.0</b>

(a) Excludes 7,078 persons who responded in the PES as Australian born and 3,418 persons who responded in the PES as born overseas, who were not matched to Census.

6 The PES used an estimation technique to adjust the weights of responding persons according to their PES-reported categories. This ensured that PES respondents not counted in the Census received the same weight adjustments as PES respondents who were counted in the Census. The Census categories were then used in producing sample-based estimates of the number of persons counted in the Census. The weighting ensured these estimates matched the actual Census counts for all benchmark categories.

## Technical Note 2 : Reliability of the PES Estimates

### TECHNICAL NOTE 2: RELIABILITY OF THE PES ESTIMATES

#### SAMPLING ERRORS ASSOCIATED WITH STATISTICS PRODUCED FROM THE PES

1 Statistics produced from the PES are subject to sampling error. Since only a sample of dwellings was included in the PES, estimates derived from the survey may differ from figures which would have been obtained if all dwellings had been included in the survey. Further, the particular sample selected for the PES was only one of a number of possible samples and each possible sample would also yield different estimates. One measure of the likely difference is given by the standard error (SE).

2 Given an estimate and the SE on that estimate, there are about two chances in three that the sample estimate will differ by less than one SE from the figure that would have been obtained if all dwellings had been included in the survey, and about nineteen

chances in twenty that the difference will be less than two SEs.

**3** The following example illustrates the use of the concept of SE. If an estimate of 2.5% has a SE of 0.1 percentage points, there are two chances in three that the figure that would have been obtained if all dwellings had been included in the sample is in the range  $2.5\% \pm (1 \times 0.1\%)$ ; i.e. between 2.4% and 2.6%, and nineteen chances in twenty that the figure is in the range  $2.5\% \pm (2 \times 0.1\%)$ ; i.e., between 2.3% and 2.7%.

## SAMPLING ERRORS ON ESTIMATES OF DIFFERENCE

**4** The sampling error on the difference between two estimates can be derived from their SEs. For the difference between two estimates  $x$  and  $y$  produced from the PES, the SE of the difference may be approximated by the following formula:

$$SE(x - y) = \sqrt{[SE(x)]^2 + [SE(y)]^2}$$

**5** This approximation will be exact for differences between estimates in different states, for greater capital city versus rest of state regions, or for differences between estimates from different Censuses. However, for estimates within the same region, there will be a negative correlation between the estimates (or rates) so that the approximation will generally underestimate the true SE.

**6** For example, if the estimates of the rate of net undercount for usual residents in Queensland greater capital city and rest of state are 2.7% and 1.4%, with SEs of 0.35 and 0.5 percentage points respectively, and using the formula above, the SE on the difference (1.3 percentage points) is:

$$SE(x - y) = \sqrt{(0.35)^2 + (0.5)^2} = 0.61 \text{ percentage points}$$

**7** Therefore there are nineteen chances in twenty that the difference between the rates of undercount for usual residents between these two regions is within the range  $1.3 \pm (2 \times 0.61)$ ; i.e., between 0.08 and 2.52 percentage points.

## NON SAMPLING ERROR

**8** The estimates of net undercount are also subject to non-sampling errors which occur in all collections, whether censuses or surveys. Examples of this kind of error include imperfections in reporting by respondents and errors made in the collection and processing of data. Every effort is made in both the Census and PES to minimise non-sampling error by careful design of forms, training and supervision of field officers and interviewers, and by using effective operating procedures. Types of non-sampling error arising from the way the PES is conducted and the way estimates are derived from the survey are discussed below.

**9** A potential weakness in the PES method is its necessary dependence on linking as a means of deciding whether or not a given person or dwelling in the PES has been counted in the Census. Despite procedures to minimise this, the difficulties associated with the linking process mean there is a risk of failing to link people who were actually included in the Census. The effect of not linking when there should have been a link would be to overstate net undercount in the Census. However, the introduction of ADL in the 2011 PES processing phase, which was used again in 2016, helped to reduce the likelihood of this type of error occurring.

**10** Nevertheless, if the variables used to establish the link are of poor quality (e.g. not stated or imputed), links are less likely to be made. To mitigate this risk, Census records that had insufficient personal identifier information, and therefore did not have a high chance of being linked to PES, were moved to the Census non-contact sector and treated in a similar fashion to late returns. This avoids a bias in the contact sector, albeit at the expense of increased sample error from the non-contact sector. For more information on the treatment of these records, see **Components of Net Undercount** on the Summary tab.

**11** While the Census and PES are conducted independently of each other, they are very similar in many respects. Thus, some weaknesses in the Census may also be shared by the PES leading to an understatement of net undercount. For example, dwellings missed by a Census Field Officer are often difficult to find and so could be missed by a PES interviewer as well. In addition, people who avoid being included in the Census may also avoid being included in the PES. The use of benchmarks in estimation helps to control for the effect of this 'correlation bias'.

## ERRORS ASSOCIATED WITH THE NON-CONTACT SECTOR

**12** The PES provides an estimate of the total number of people who should have been counted in the Census non-contact sector (i.e. late returns, imputed persons in non-responding dwellings, and persons with insufficient personal identifier information on their Census form).

**13** PES estimates of the population in the non-contact sector have relatively high sampling errors, mainly because Census person counts for this sector were not available to use as a weighting benchmark, but also due to the small sample size (as there were relatively few Census non-contact dwellings selected by chance in the PES sample).

**14** This lack of Census person counts also means that, while the dwelling weights used for the non-contact sector were estimated from the sector itself, the adjustments applied to provide final person weights use the information observed in the contact sector. This is a potential source of non-sampling error, as is any bias arising from peculiarities of the non-respondents in this sector. Both these sources of non-sampling error are expected to be small, compared with the sampling error of the population estimate for the non-contact sector.

## Glossary



## **GLOSSARY**

### **Aboriginal and Torres Strait Islander Peoples**

People who identified themselves, or were identified by another household member, as being of Aboriginal origin, Torres Strait Islander origin, or both.

### **Australian Statistical Geography Standard (ASGS)**

The ASGS brings together within one framework all the regions on which the ABS publishes statistics. This is used by the ABS for collection and dissemination of geographic statistics and provides the context for understanding and interpreting the geographic context of statistics published by the ABS. It is a hierarchically structured classification split into two broad groups, ABS structures and Non-ABS structures, to satisfy different statistical purposes.

For more information, see the ABS publication Australian Statistical Geography Standard (ASGS): Volume 1 – Main Structure and Greater Capital City Statistical Areas, July 2016 (cat. no. 1270.0.55.001).

### **Automated Data Linking (ADL)**

ADL is an automated linking processes used to determine possible links between Census and PES data, before any clerical matching process has begun. It employs a probabilistic linking method that uses a range of personal and addresses characteristics to evaluate the likelihood that a PES and Census record pertain to the same individual. ADL is run via the Freely Extensible Biomedical Record Linking (FEBRL) software.

### **Benchmark Category**

A benchmark category is a PES estimate (either dwelling or person based) of the Census count in a category (e.g. Country of birth) based upon the Census response to that category, regardless of the PES response. For example, if a Census record has Australia as the country of birth, but this is recorded as Born Overseas in the PES, the record will be in the Australia benchmark category.

### **Census Count**

The Census counts people where they were located on Census night and this count of the population is referred to as the place of enumeration count. A count of the population based on their usual residence is also available. The Census count includes imputed persons for non-responding dwellings. However, some categories (such as Indigenous status and Country of birth) do not include any imputed persons, as Census assigns a not stated value to all the imputed person records.

### **Computer Assisted Interviewing (CAI)**

CAI is a method of data collection whereby responses are recorded directly into an electronic questionnaire on a notebook computer.

### **Contact Sector**

The Census contact sector comprises: persons in dwellings that were deemed occupied on Census night and from which a form was received before the commencement of PES enumeration; persons in occupied dwellings that were entirely missed by Census; and persons missed by Census because their dwellings were mistakenly deemed unoccupied on Census night.

### **Discrete Community**

A Discrete Community is a geographic location bounded by physical or legal boundaries, which is inhabited or intended to be inhabited predominantly by Aboriginal or Torres Strait Islander peoples, with housing or infrastructure that is managed on a community basis.

### **Dwelling**

A dwelling is a building or structure in which people may live. Examples of dwellings include: houses, flats, caravans, tents, humpies and houseboats. Houses under construction, derelict houses, vacant tents, or unoccupied converted garages, are not counted as dwellings in the Census. Dwellings are categorised as either private or non-private.

### **Dwelling Link Rating (DLR)**

The DLR is a numeric indicator of the strength of a PES-Census dwelling link, used to rank the links for preference in later processing. These are derived for PES dwellings primarily by combining the Person Link Rating (PLR) of individuals in the dwelling with the addition of some other criteria.

### **Estimated Resident Population (ERP)**

ERP is the official measure of the population of Australia based on the concept of usual residence. It refers to all people, regardless of nationality, citizenship or legal status, who usually live in Australia, with the exception of foreign diplomatic personnel and their families. It includes usual residents who are overseas for less than 12 months over a 16-month period. It excludes overseas visitors who are in Australia for less than 12 months over a 16-month period.

### **Freely Extensible Biomedical Record Linking (FEBRL)**

FEBRL is the software application used to run the Automated Data Linking (ADL) process in the 2011 and 2016 PES.

### **General Population Sample**

The term General Population refers to the PES private dwelling sample, excluding those dwellings selected from the Discrete Community sample.

### **Greater Capital City/Rest of State/Territory**

Greater Capital City Statistical Areas (GCCSA) are geographical areas built from Statistical Areas Level 4 (SA4), designed to represent the functional extent of each of the eight State and Territory capital cities.

There are 16 spatial GCCSA regions covering the whole of Australia without gaps or overlaps. These consist of 8 regions representing each of the Australian State and Territory capital cities, and 8 regions covering the rest of each State and the Northern Territory. The ACT only has one GCCSA region for the entire Territory. The category of Other Territories in GCCSA includes the Other Territories of Jervis Bay, Christmas Island, Cocos (Keeling) Island and Norfolk Island.

For more information, see the ABS publication Australian Statistical Geography Standard (ASGS): Volume 1 – Main Structure and Greater Capital City Statistical Areas, July 2016 (cat. no. 1270.0.55.001).

### **Hot-deck Imputation**

Hot-deck imputation is the primary imputation method used for the 2016 Census. The method involves locating a donor record and copying the relevant responses to the record requiring imputation. The donor record will have similar characteristics and must also have the required variable(s) stated. In addition, the donor record will be located geographically as close as possible to the location of the record to be imputed.

### **Imputation**

Imputation is a statistical process for predicting values where no response was provided to a question and a response could not be derived. Census imputes persons into non-responding dwellings, using a hot-deck imputation method, and also imputes some missing values (e.g. Age and Sex) for responding persons who left these fields blank.

### **Indigenous Community Frame (ICF)**

The ICF is a listing of all Discrete Communities from which the sample of communities for the 2016 PES was selected. The selection unit on the ICF is a community set, containing one main community and zero or more outstations. If a community set is selected, a random selection of dwellings from the main community, and all the dwellings in some outstations, will be enumerated in the survey.

### **Late Return**

A late return is a Census form which was returned after the start of PES enumeration.

### **Match and Search System (MSS)**

The MSS is the main PES clerical review facility, which allows processors to search, view, compare, and record matches between PES and Census records.

### **Net Undercount**

Net Undercount is the difference between the PES estimate of the number of people who should have been counted in the Census and the actual Census count (including persons imputed into non-responding dwellings). The estimated net undercount for a category of person is the net result of undercount, overcount, differences in classification between the PES and Census (e.g. Age, Sex, Indigenous status) and imputation error in the Census.

### **Non-contact Sector**

The Census non-contact sector comprises: persons in dwellings that were deemed occupied on Census night, from which no Census form was received (imputed dwellings); persons in dwellings whose Census form was received after the commencement of PES enumeration (late returns); and persons with insufficient personal identifier information on their Census form.

### **Non-private Dwelling (NPD)**

An NPD is an establishment which provides a communal type of accommodation, such as a hotel, motel, hospital or other institution. NPDs were not included in the 2016 PES sample.

### **Outstation (or homeland)**

An Outstation is a discrete Aboriginal and Torres Strait Islander Community that is administered by, or linked to, an organisation such as a resources agency or larger parent discrete Aboriginal and Torres Strait Islander Community for the provision and maintenance of services.

### **Person Link Rating (PLR)**

The PLR is a numeric indicator of the strength of a person link. This is generated through the Collect, Analyse, Reduce, De-duplicate and Systematise (CARDS) process and used to stratify ADL links by quality. It is used throughout matching processing and contributes to the DLR of the PES dwelling in which the person was enumerated.

## **PES Address**

A PES Address is any address associated with a PES respondent; includes: the Enumeration address (the address at which the PES interview took place), and Search addresses (any other address at which the respondent might have been included on a Census form).

## **PES Population Estimate**

The PES population estimate is an estimate (based on PES and Census data) of the number of people who should have been counted in the Census.

## **Private Dwelling**

A private dwelling is a residential structure which is self-contained, owned or rented by the occupants, and intended solely for residential use. A private dwelling may be a flat, part of a house, or even a room, but can also be a house attached to, or rooms above, shops or offices.

## **Sampling error**

Sampling error occurs because a sample, rather than the entire population, is surveyed. The standard error gives a measure of the likely difference resulting from not including all dwellings in the survey. There are about two chances in three that a sample estimate will differ by less than one standard error from the figure that would have been obtained if all dwellings had been included in the survey, and about nineteen chances in twenty that the difference will be less than two standard errors.

## **Search Address**

A search address is an address (other than the PES enumeration address) where a person reported they were staying on Census night or where a person may have been included on a Census form (including non-reported addresses found by ADL). PES processing uses these addresses to locate a Census form in order to determine the number of times (if any) a person enumerated in the PES was included on a Census form.

## **Standard Error (SE)**

The SE is a measure of the likely difference between the true value and the estimate.

## **Statistical Areas 1, 2, 3 & 4**

Statistical areas are the geographic areas under the ASGS geographic classification used in the PES. Each area is built from aggregates of the lower area and cover Australia without gaps or overlap.

For more information, see the ABS publication Australian Statistical Geography Standard (ASGS): Volume 1 – Main Structure and Greater Capital City Statistical Areas, July 2016 (cat. no. 1270.0.55.001).

## **Undercount Adjustment Factor**

The undercount adjustment factor is the ratio of the PES population estimate to the Census count. This factor can be applied to the Census counts to indicate how many people should have been counted in the Census for that category.

## **Usual Residence**

In the Census, the usual residence for a resident of Australia is defined as the place where they have lived, or where they intend to live, for 6 months or more, in the Census year.

# **Abbreviations**

## **ABBREVIATIONS**

ABS	Australian Bureau of Statistics
ACT	Australian Capital Territory
ADL	Automated Data Linking
ASGS	Australian Statistical Geography Standard
CFA	Census Field Area
ERP	Estimated Resident Population
MSS	Match and Search System
NPD	Non-private Dwelling
NSW	New South Wales
NT	Northern Territory
PES	Census Post Enumeration Survey
PREG	Prediction Regression
QA	Quality Assurance
Qld	Queensland
SA	South Australia
SA1	Statistical Area Level 1
SE	standard error

## Quality Declaration - Summary

### QUALITY DECLARATION - SUMMARY

#### INSTITUTIONAL ENVIRONMENT

For information on the institutional environment of the Australian Bureau of Statistics (ABS), including the legislative obligations of the ABS, financing and governance arrangements, and mechanisms for scrutiny of ABS operations, please see ABS Institutional Environment.

#### RELEVANCE

The Census of Population and Housing is the largest statistical collection undertaken by the Australian Bureau of Statistics (ABS) and one of the most important. Its objective is to accurately measure the number of people in Australia on Census night, their characteristics, and the dwellings in which they live. Due to its size and complexity, however, it is inevitable that some people will be missed from the Census, some will be counted more than once, and some will be counted in error.

The Census Post Enumeration Survey (PES) is a household survey conducted by the ABS shortly after each Census, in order to provide an independent measure of Census coverage. The PES results are used to determine how many people should have been counted in the Census, how many people were missed (undercount), how many were counted more than once (overcount) and how many were counted in error. It also provides information on the characteristics of those in the population who have been missed or overcounted. Together with the Census counts, these components determine the Census net undercount.

PES estimates of net undercount augment the Census counts for the purpose of deriving estimated resident population (ERP) for Australia and its states and territories.

#### TIMELINESS

The **Census and Statistics Act 1905** requires the Australian Statistician to conduct a Census every five years. The PES is also conducted every five years, shortly after each Census.

The 2016 PES has a single data release, made available on the ABS website on 27 June 2017. These results are published in conjunction with both Census results and rebased ERP.

#### ACCURACY

The 2016 PES was designed to provide reliable estimates at the national level and for each state and territory.

Dwellings in each state and territory were selected at random using a multi-stage area sample. The sample for the 2016 PES consisted of approximately 54,000 private dwellings from the ABS private dwelling and Indigenous Community frame.

Estimates in this publication are subject to sampling and non-sampling errors. Sampling error is the error associated with taking a sample of dwellings rather than going to all dwellings in Australia. In this publication the sampling error is measured by the standard error (SE) which is expressed as a percentage of the estimate. Non-sampling errors can occur in any data collection and sources include non-response, errors in reporting by respondents or recording answers by interviewers, and errors in coding or processing of data. Every effort is made to minimise reporting error by the careful design of questionnaires, intensive training and supervision of interviewers, and efficient data processing procedures.

#### COHERENCE

It is important for PES data to be comparable and compatible with previous PESs and also with other data produced by the ABS and wider community. The ABS, and the PES, uses Australian standard classifications, where available and appropriate, to provide data comparability across statistical collections. These include, for example, standards for geographic areas.

PES publications for 1991, 2001, 2006 and 2011 are also available on the ABS website.

#### INTERPRETABILITY

This publication contains tables and a summary of findings to assist with the interpretation of the results of the survey. Detailed Explanatory Notes, Technical Notes on Data quality and a Glossary are also included, providing information on the terminology, classifications and other technical aspects associated with these statistics.

#### ACCESSIBILITY

If the information you require is not available as a standard product, then ABS Consultancy Services can help you with customised services to suit your needs. For inquiries contact the National Information and Referral Service on 1300 135 070. Alternatively, please email [client.services@abs.gov.au](mailto:client.services@abs.gov.au)

